Chapter 6 Dataset Codebook

# ch6\_health\_hazards Dataset

This dataset contains Census tract level FIPS codes for 210 Census tracts in St. Louis, Missouri, and the environmental health hazards index for each tract. The data come from National Air Toxics Assessment (NATA) conducted by the US Environmental Protection Agency (EPA) and was accessed through the eGIS section of hud.gov. The environmental health hazards index is calculated for all Census tracts in the United States, and it measures potential exposure to harmful toxins in the environment. The index takes into account carcinogenic, respiratory, and neurologic hazards, and indicates the national percentile rank of the Census tract. The lowest value is 0, indicating maximum exposure to toxins harmful to health, and the highest value is 100, the best environmental quality and least exposure to toxins.

Reference: http://egis.hud.opendata.arcgis.com/datasets/53a856bef6f24356abee30653399e94a\_0

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| census\_tract | Eleven digit number | Census tract |
| health\_hazards\_index | Numeric | Environmental health hazards index,  theoretical minimum = 0 (maximum exposure to toxins), theoretical maximum = 100 (least exposure to toxins, best environmental quality) |

# ch6\_youth\_drinking Dataset

This dataset comes from the Youth Risk Behavior Surveillance System (YRBSS) 2013, which is run by the Centers for Disease Control and Prevention (CDC). There are observations from 200 students who reported missing school because they felt unsafe at school or on their way to or from school on at least 1 day during the 30 days before the survey. There is an indicator for whether or not the student drank at least one drink of alcohol on at least 1 day during the 30 days before the survey.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| missed\_school | 1 = “Yes” | Indicator for students who did not go to school because they felt unsafe at school or on their way to or from school on at least 1 day during the 30 days before the survey; all students in dataset have a value of 1 |
| alcoholic\_drink | 1 = “Yes”  0 = “No” | Indicator for students who currently drank at least one drink of alcohol on at least 1 day during the 30 days before the  survey |

# ch6\_seat\_belt Dataset

The data come from 2014 Behavioral Risk Factor Surveillance System (BRFSS) operated by the Centers for Disease Control and Prevention (CDC). The dataset includes a unique identifier for all 459 subjects in the dataset. Each subject was asked,” How often do you use seat belts when you drive or ride in a car?” and those who answered “Always” received a value of “1” for wear\_seat\_belt. Those who reported not always wearing a seat belt were coded as “0.”

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric 1 – 459 | Identifier |
| wear\_seat\_belt | 1 = “Yes”  0 = “No | Indicator for those who report always wearing a seat belt when driving or riding in cars |

# ch6\_birth\_weight Dataset

The dataset includes data from 247 male babies sampled in the 2013 – 2014 National Health and Nutrition Examination Survey (NHANES). Each subject has a unique identifier and a value for birth weight in pounds.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric 1 – 247 | Identifier |
| birth\_weight | Numeric | Birth weight in pounds |

# ch6\_arrests Dataset

The dataset compiled by ProPublica includes data from 1,332 females arrested in Broward County, Florida, from 2013 – 2014. Each subject has a unique identifier, number of prior crimes committed (criminal history), and an indicator for those with recidivism within two years. Recidivism is defined as a new arrest within two years, according to Northpointe’s practitioners guide. The guide says the recidivism score predicts “a new misdemeanor or felony offense within two years of the COMPAS administration date.” The recidivism variable sought to determine if a person had been charged with a new crime after the crime for which they were COMPAS screened. Traffic tickets and some municipal ordinance violations did not count as recidivism. People arrested for failing to appear at court hearings or later charged with a crime before their COMPAS screening were not counted as recidivists. The database of COMPAS scores was supplemented with a profile of each person’s criminal history, which contained information from before and after they were scored using public criminal records from the Broward County Clerk’s Office website through April 1, 2016.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric 10 -11001 | Identifier |
| priors\_count | Numeric | Number of prior crimes committed |
| two\_year\_recid | 1 = “Yes”  0 = “No | Indicator for those with a new arrest within two years of being scored by COMPASS |